A Semantic Map for Evaluating Creativity

Frank van der Velde[1,2], Roger A. Wolf[3], Martin Schmettow[1] and Deniece S. Nazareth[1]
[1]Cognitive Psychology and Ergonomics (CPE-BMS), CTIT, University of Twente,
Drienerlolaan 5, 7522 NB Enschede, The Netherlands
[2]IOP Leiden University, The Netherlands
[3]Saxion University of Applied Sciences,
Handelskade 75, 7417 DH Deventer, The Netherlands
f.vandervelde@utwente.nl

Abstract

We present a semantic map of words related with creativity. The aim is to empirically derive terms which can be used to rate processes or products of computational creativity. The words in the map are based on association studies performed by human subjects and augmented with words derived from the literature (based on human raters). The words are used in a card sorting study to investigate the way they are categorized by human subjects. The results are arranged in a heat map of word relations based on a hierarchical cluster analysis. The cluster analysis and a principal component analysis provide a set of five to six clusters of items related to each other, and as clusters related to creativity. These clusters could form a basis for scales used to rate aspects of computational creativity.

Introduction

In his *Principles of Psychology*, published in 1890, William James introduced his definition of 'attention' as follows: "Everyone knows what attention is". Yet, debates on the distinctive features of attention continue up to the present day.

Perhaps a similar situation could be found with the notion of 'creativity'. In some way, 'everyone knows what creativity is'. But it is non-trivial to find methods by which creativity can be evaluated. Yet, when we investigate creativity, either in humans or as achievements of computational systems, we need some way to evaluate creativity. For example, we need a measure of creativity to distinguish between brain states in the neuroscientific investigation of creativity (Fink and Benedek, 2014). We also need it to assess the products of computational systems as creative or not. Indeed, the question of how computational creativity can be evaluated has been described as one of the 'Grand Challenges' of computational creativity research (Cardoso, Veale and Wiggins, 2009).

Definitions of creativity have been presented in the literature. For example, "creativity is commonly defined as the ability to produce work that is both novel (original, unique) and useful" (Fink and Benedek, 2014, p. 111). Similar characteristics are novelty and useful or value (Amabile, 1996; Hennessey and Amabile, 2010), typicality, novelty and quality (Ritchie, 2007), novelty, value, and unexpectedness or surprise (Grace and Maher, 2014), and skill, imagination and appreciation (Colton, 2008).

Each of these qualifications may capture aspects of creativity. But when they are used as criteria for the evaluation of creativity by human raters, as in the evaluations of processes or products of computational creativity, we need to validate their relation with the notion of creativity. In this context it is important to realize that an assessment (rating) performed by humans is an empirical investigation (behavioral experiment), whether or not the raters are experts or arbitrary people, and the rating scales used are instruments of measurement, which need to be validated. For this, it is not sufficient to argue that the rating scales are based on some kind of definition (no matter how sound the definition may appear to be).

Recently, Jordanous (2012a, 2012b, 2014) investigated the question of how creativity of computational creativity systems is and should be evaluated. Based on an analysis of the evaluation of creativity in the scientific literature related to computational creativity, she found that evaluation ratings (if performed at all) were based on criteria set up by the researchers themselves (or by other researchers in the literature).

To achieve a more empirical basis for rating computational creativity (i.e., not just derived from the subjective acceptance by researchers), Jordanous (2012a,b) used a statistical analysis by comparing word frequencies in scientific articles related to the study of computational creativity with word frequencies in scientific articles related to other topics. An analysis of this kind is based on the assumption that the meaning of words is related to the context in which the words are used. In particular, the meaning of a word (or aspects of it) can be determined by finding other words that co-occur with it statistically more often than can be expected on the basis on chance (Landauer and Dumais, 1997).

Based on her analysis, Jordanous (2012a,b) derived a set of 694 terms that occurred more frequently in the scientific literature related to computational creativity comparted to other, non-related, scientific articles. On the basis of these words, she derived 14 dimensions on which creativity could be evaluated.

Here, we investigate the empirical basis for rating (computational) creativity based on empirical (behavioural) studies with human subjects. After all, ratings of creativity are conducted by human subjects, so we could also probe human subjects for the basis of these rating scales. Our aim is to arrive at a 'semantic map' of terms related to the notion creativity, which can be used to derive and compare rating scales for creativity.

To arrive at this semantic map, we conducted a study in which human subjects were asked to provide terms associated with creativity. Next, the terms associated with creativity were used in a 'reverse' association study, to see whether terms like 'creativity' are in turn associated with these terms. Then, a selected set of words based on both association studies was used in a card sorting study with human subjects. The words used in our card sorting study were augmented with a selected subset of the 694 words related to creativity based on the analysis of Jordanous (2012a,b). A card sorting study provides information about how a set of words are categorized by human subjects. Using the words based on our association studies, this in turn provides a prototype for a semantic map related to creativity.

The remainder of this article is structured as follows. First, we outline how a set of words was derived as the basis for the semantic map. Then, we present and discuss the card sort study used to derive the semantic map. Next, the prototype of the semantic map based on the card sorting study is presented and discussed. Finally, we present the conclusions and briefly discuss future work.

Word associations with creativity

As introduced above, we conducted two word association studies. Word associations are used as a technique in experimental psychology, for example to obtain controlled stimulus material (Nelson, McEvoy & Schreiber, 2004).

In an association study, a target word is given and subjects are asked to produce words associated with the target. In a free association study a subject can give an unlimited number of associated words. In a restricted or discrete association study, the number of association words is restricted beforehand (in case of a discrete study, only one associated word can be given). A problem with a free association study is the occurrence of a chain of associations, in which (new) associated words are given not because they are associated with the target word but instead are associated with a previously given associated word. We therefore used a restricted and a discrete association study.

The aim of our first association study was to derive a set of terms associated with the word 'creativity'. For this, we conducted a restricted association study. In this study, 36 subjects between the age of 18 and 52 (29 Dutch and 7 German) were asked to give at most three terms associated with the word 'creativity' (either in Dutch or German). From this list three human raters selected a list of words on which they all agreed as words associated with creativity. This resulted in a set of 58 words.

We augmented this list by a selection of words based on the set of words derived by Jordanous (2012a,b). She analyzed two corpora of texts: one consisting of scientific articles related to the study of creativity and one consisting of scientific articles not related to the study of creativity. A statistical analysis revealed a set of 694 terms that occurred statistically more frequently in the scientific articles related to the study of creativity. In our study, this set was reviewed by three human raters. They each selected words from this set that in their view were associated with creativity. The words on which all three raters agreed were included in the set of words associated with creativity. This procedure resulted in an initial list of 32 words based on the list provided by Jordanous (2012a,b).

The list of 58 words obtained in our first association study included 10 words from the list of Jordanous (2012a,b) selected by the three human raters (see above). The list of 58 words included another eight words from the list of Jordanous (2012a,b) which were not selected by the three human raters.

In this way, we obtained a list of 80 words to be used in our second association study. In this list of 80 words, 22 words derived exclusively from the list of Jordanous (2012a,b), in the manner outlined above; 40 words were derived exclusively from the list provided by human subjects in our first association study; 18 words co-occurred in the list of Jordanous and in the human subject list obtained in our first association study.

In our second association study we used the list of 80 words obtained in our first association study, augmented with the words selected from Jordanous (2012a,b), to conduct a 'backward' (or reverse) discrete association study. That is, for each of these 80 words human subjects were asked to provide one term associated with that word. The list of words was presented in a randomized order to prevent priming effects. A subject sat in front of a screen and a keyboard in an isolated cubicle. One word at a time appeared on the screen. The subject then used a keyboard to type the answer. After that, a new word appeared. The subjects consisted of 50 students between age 19 and 27. None of them participated in the first part of the study. There were 29 Dutch and 21 German participants from whom 24 were men and 26 women. There were 25 technical students, 22 social studies students and 3 art students.

The first aim of our second association study was to obtain 'reversed' associations to the words associated with creativity (the list of 80 words outlined above). In particular, to see whether words like 'to create', 'creative' or 'creativity' are in turn associated with the words associated with the word 'creativity'. A second aim of this study was to see whether words in the list of 80 words are associated with each other.

A subset of the list of 80 words gave a 'creativity' word ("creativity", "creative" or "to create") as a (reversed) association in our second association study. In this subset, 55% of the words came from the human list derived in our first association study, 28% from the list provided by Jordanous (2012a,b) and 17% from both lists. However, the whole list of ‘reversed’ associated words obtained in our second association study was used as one of the lists on which the words for the card sorting study were based, in the manner outlined below.

Card sorting study

The list of words obtained in our first association study (augmented with words from the list of Jordanous) and the list of words obtained in our second association study were used to select the words for the card sorting study.

	Words used in card sorting	Source: H (Human); J (Jordanous); B (Both)
1.	Different	H
2.	Artistic	B
3.	(To) Knit	H
4.	Extraordinary	B
5.	(To) Think	B
6.	Imagination	B
7.	Thought	J
8.	Poem	B
9.	Mind	H
10.	Feeling	H
11.	Craftsmanship	H
12.	Idea	B
13.	Hunch	B
14.	Innovation	J
15.	Inspiration	B
16.	Intelligence	J
17.	Knowledge	J
18.	Colours	H
19.	(To) Craft	H
20.	Art	H
21.	(To) Make	H
22.	Difficult	H
23.	Music	B
24.	Novelty	B
25.	Unconventional	B
26.	(To) Design	H
27.	Original	B
28.	Passion	H
29.	Planning	H
30.	Process	J
31.	Painter	B
32.	(To) Play	B
33.	Spontaneity	B
34.	Talent	J
35.	(To) Draw	H
36.	Invent	J
37.	Unique	H
38.	Skill	J
39.	Renewing	H
40.	Realise	H
41.	Resourceful	H
42.	Happy	H

Figure 1. List of words used in the card sorting study

The selection was based on three conditions:

Firstly, a word had to appear in both lists of words. Thus, a word is considered to be strongly associated with creativity if that word is both directly and indirectly (reversely) associated with creativity. Direct association entails that the word is associated with creativity (more specifically, the word belongs to the word list of our first association study, augmented with words from Jordanous, 2012a,b). Indirect association entails that the word is associated with a word that is in turn associated with creativity (more specifically, the word belongs to the list of words obtained in our second association study).

Secondly, a word had to appear more than once as an answer in our second association study (to avoid the use of idiosyncratic words in the card sorting study).

Thirdly, the word could not be the word “creative” or any derivative of that base word, because the aim of this card sorting study was to investigate the internal semantic structure of the words strongly associated with creativity without interference from the base word “creativity” itself.

In all, 42 words were selected for the card sorting study. In the study 40 Dutch participants took part. They did not participate in any of the previous studies. Figure 1 presents the words used in the card sorting study and the source (lists) on which they are based. That is, the source consists of the list derived from our association studies (H, 19 words); the list of Jordanous (2012a,b) (J, 8 words); or both lists (B, 15 words).

Card sorting can be used to evaluate how people organize a set of items (Harloff and Coxon, 2006). Figure 2 illustrates a card sorting study with the following set of words: *keyboard, printer, mouse, cat, dog*.

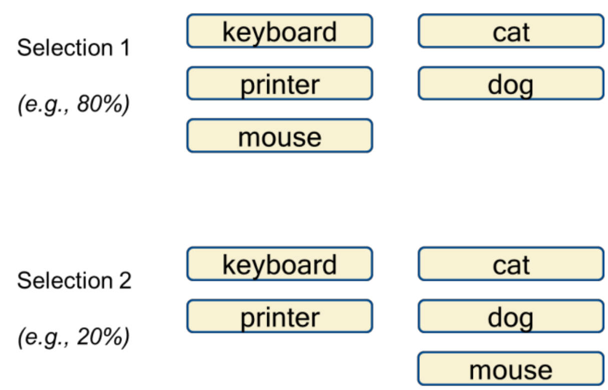

Figure 2: Example of a card sorting study

In a card sorting study, these words are printed on cards and subjects are asked to group these cards into categories[1]. If, in their view, a word cannot be placed in a category, it forms a category on its own. All words have to be selected in this way. The set of words in figure 2 could, for example, be grouped as {*keyboard, printer, mouse*} and {*cat*, *dog*} (selection 1) or as {*keyboard, printer*} and {*mouse, cat, dog*} (selection 2). The number of times (percentage) a particular categorization is chosen determines the (relative) strength of that categorization.

[1] One can also use an online version of a card sorting study. For an example, see https://conceptcodify.com/studies/jfvi9n5751vue9bn/via/demo_use_only_not_recording/

The results of the card sorting study with our set of 42 word associated with creativity were analyzed with a Hierarchical Cluster Analysis (HCA), using the statistical programming environment R (Salmoni, 2012). The HCA technique (Coxon, 1999) selects the two highest associated words (i.e., that most often occur together in a card sorted group) and replaces them with a single item. The associations of this item with the other words are the average of those of the two words forming the item. Continuing in this way, a hierarchical cluster can be obtained of the results of the card sorting study.

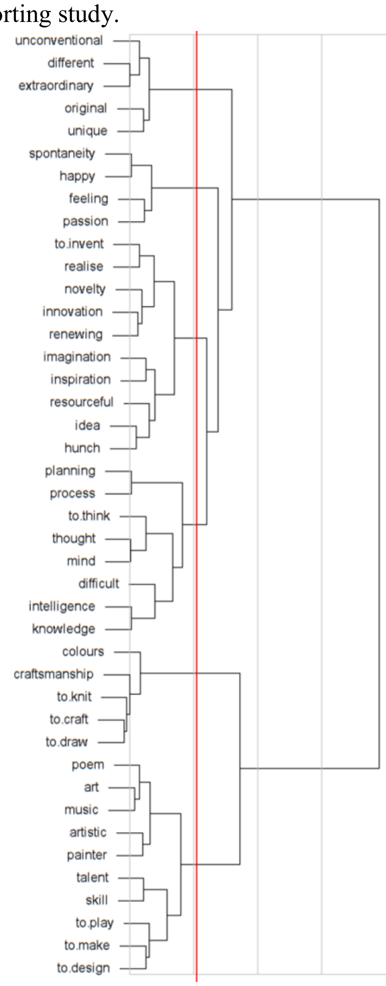

Figure 3: Hierarchical clustering of the 42 words used in the card sorting study of terms associated with creativity

The results of the HCA on the card sorting data are presented in Figure 3. The hierarchical cluster structure provided by the HCA starts with clusters of one or two words at the left and ends with two overall clusters at the right. The horizontal distances in figure 3 provide a measure of (relative) distance between clusters and subclusters. Short distances between subclusters (as between the first layer of clusters at the left of the hierarchy) suggest that they essentially form a larger subcluster. Visual inspection of the HCA suggests that a set of subclusters to the left of the red line might provide information about a meaningful classification of the words related to creativity, because the distances within these subclusters are relatively short compared to the distances between the subclusters.

Figure 4 presents a set of basic clusters of terms associated with creativity, based on the HCA presented in figure 3. They are selected (as indicated by the red line), by using the same distance from the basis as a selection measure. A basis for the selection is the observation that item-distances between clusters are substantially larger than item-distances within clusters.

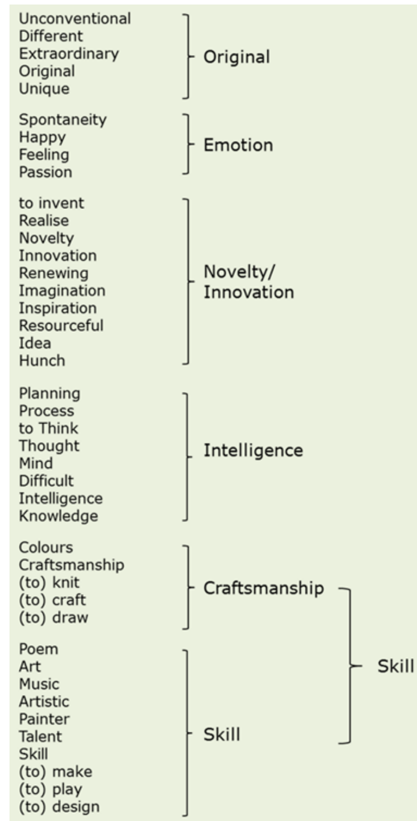

Figure 4: Tentative clusters related to creativity

Figure 4 presents six clusters and tentative cluster names. Perhaps the last two clusters could be combined into one, given that the item-distances between these clusters and the other clusters are the largest distances of the hierarchy in figure 3. This would provide the following five main clusters of items associated with the concept creativity:

- Original (originality)
- Emotion (emotional value)
- Novelty / innovation (innovative)
- Intelligence
- Skill (ability)

Before discussing these clusters we present and discuss a further analysis of the data based on the 'heat map' presentation of the results from the card sorting study.

Heat map of card sorting results

The results of the card sorting study can also be represented in a heat map, in which the color indicates the strength of the association between two terms.

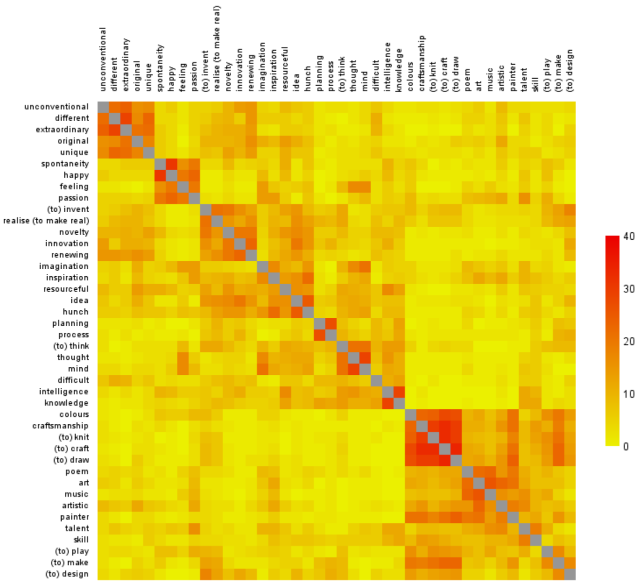

Figure 5: Heat map presentation of the card sorting results

Figure 5 presents the heat map based on the results of the card sorting study. The rows and columns in the heat map represent the words used in the card sorting study (figure 1). The words in the heat map are arranged in the order of the HCA analysis presented in figure 3. In this way, the heat map forms a matrix. The color in each matrix cell represents the number of times the row and column word corresponding to the cell belonged to the same group in the card sorting study. Given that 40 subjects participated in the study, this number can vary between 0 and 40. The heat map presents this number in terms of a color, varying from light yellow (0) to deep red (40). In the data, the lowest number was 0 and the highest number was 34. The heat map is symmetric because the words used in the card sorting study are represented as rows and as columns. For this reason, the diagonal in the heat map does not represent data from the card sorting study.

It is clear that the squares that form groups of words are related to the clusters in figure 3 (which results from the fact that the words in the heat map are arranged in the order of the HCA analysis presented in figure 3). For example, in the top left corner there is a 5x5 square that is much more red (darker) than the yellow around it. This 5x5 square belongs to a group of five words: *unconventional, different, extraordinary, original* and *unique*. If we wanted to label this group with one name, it could be 'original', as indicated by the cluster name in figure 4. Original is often referred to in the literature as a characteristic of creativity (e.g., Hennessey and Amabile, 2010). Also, in the right corner at the bottom we see a large group that is relatively distinct from the rest. This is the group that we labeled as 'skill' in figure 4. This group comprises a smaller 'skill' group and a 'craftsmanship' group in figure 4 (the 'craftsmanship' group stands out within the larger 'skill' group in the heat map). 'Skill' has also been related to creativity in the literature (e.g., Colton, 2008).

Yet, although the HCA structure in figure 3 and the heat map in figure 5 are based on the same data, they reveal different aspects of the semantic map based on the card sorting study of terms associated with creativity.

The HCA structure shows a metric within and between the clusters of terms related to creativity. The metric is given by the (vertical) distance that needs to be travelled in going from one word to another. So, for example, the distance between *unconventional* and *innovation* is shorter than that between *unconventional* and *skill*. This metric is not directly revealed in the heat map.

But the heat map shows that a word that belongs to a group can also be associated to words outside that group. For example, *unconventional* belongs to the 5 by 5 group referred to above, but it also has some association strength with *renewing*. These outside associations are not directly revealed by the HCA structure, due to the forced choice procedure on which the structure is based. In this way, the HCA analysis seems to miss the more global structure that is present in the results (and thus in the heat map). To analyze this more global structure, we analyzed the data in the heat map using a Principal Component Analysis (PCA).

PCA analysis of the card sorting results

A Principal Component Analysis (PCA) of a set of data reveals the orientations (axes) along which most of the variance in the data is found (Jolliffe, 1986; Jackson, 1991). These are referred to as the Principal Components (PCs). Starting with a covariance or correlation matrix of the data, a PCA analyses the matrix in terms of its eigenvalues and eigenvectors. The highest eigenvalue corresponds to the PC along which most of the variance in the data is found. The second eigenvalue then reveals the PC along which most of the remaining variance is found. This process continues until all of the variance in the data is accounted for. Because the eigenvectors are orthogonal, a PCA shows independent sources of variance in the data.

A PCA starts with a covariance or correlation matrix of the data. For this we used the data underlying the heat map expressed in decimal fractions (based on the maximum possible score of 40). For the diagonal values we used the score 1.0 based on the assumption that a word is maximally related to itself.

One of the advantages of a PCA is that it allows a reduction of the dimensions underlying the data, by taking into account only the PCs with the highest eigenvalues.

Figure 6 presents a graph of the eigenvalues of the heat map data in descending order. This is also known as a scree graph or scree plot (Jolliffe, 1986). A rule is to use only the eigenvalues presented by the scree plot in the section before the plot levels off. In this case that would result in representing the data based on PCs corresponding to the five highest eigenvalues (all > 2).

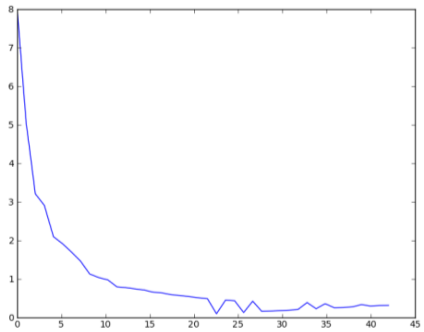

Figure 6: Scree plot of the eigenvalues in the PCA analysis of the heat map

A PCA gives the PCs of the highest variance in the data, but it does not provide an interpretation of a PC (Jackson, 1991). Looking at the heat map, however, it is clear that a substantial variance in the data results from the difference between high (red) and low (yellow) values. For a word, this difference corresponds to belonging to a subcluster (such as represented in figure 4) or not. It would seem that the first eigenvalue captures this source of variance. However, every word has both high and low values in the heat map, so this source of variance does not reveal much about the ways words belong to difference groups. Furthermore, when the values in the analyzed matrix are all positive, the coefficients of the first PC (eigenvector) are all of the same sign (Jackson, 1991).

Therefore, in figure 7 we present the words in the heat map in terms of the PCs given by the eigenvalues of the PCA of the heat map, starting with the second highest eigenvalue. The PCs are all uncorrelated, but the coefficients of a PC (eigenvector) can be correlated. These correlations are in particular affected by the signs of the coefficients (Jackson, 1991). Therefore, we group words by the signs of their coefficients for a PC. The groupings are presented in figure 7, in terms of the second to the fifth PC with the highest eigenvalues, in descending order. In figure 7, the signs of the coefficients of PC 3 to 5 are represented by the letters **P** and **N**, to indicate that different groups could have the same sign on that PC.

Figure 7 shows that the second PC (eigenvalue) separates the words in the heat map into two groups. We arranged the words in figure 7 in the manner as they appear based on 5 eigenvalues. This results in a word order (partly) different from the one found in figures 3, 4 and 5. However, it is clear that the two groups selected by the second eigenvalue in figure 7 correspond to the two largest clusters in figure 3. Thus, the first separation in the heat map (capturing most of the variance after the first eigenvalue) is between the large 'skill' cluster in figure 4 and the other words (also illustrated with the difference between the large red-like square in the bottom right corner of the heat map and the other words).

In figure 4 we selected five groups of words based on the HCA, with the 'craftsmanship' and 'skill' groups as one. In figure 7, the first four PCs also give five groups if we take the 'craftsmanship' and 'skill' groups as one. A comparison between both groupings reveals that they are quite compatible, although a few noticeable differences appear. The 'original' group in figure 4 is maintained in figure 7, with the addition of the word *renewing*, which at face value seems to be related with these words. The 'emotion' group in figure 4 is maintained as well, with the addition of *imagination* and *inspiration* (which split off with 5 PCs). So, 'emotion' may not be the correct label for this group.

Eigenvalue 2	Eigenvalue 3		Eigenvalue 4		Eigenvalue 5	
Innovation	Innovation	P	Innovation	P	Innovation	P
Idea	Idea		Idea		Idea	
Planning	Planning		Planning		Planning	
Process	Process		Process		Process	
Difficult	Difficult		Difficult		Difficult	
to invent	to invent		to invent		to invent	N
Realise	Realise		Realise		Realise	
Novelty	Novelty		Novelty		Novelty	
Original	Original		Original	N	Original	P
Unconventional	Unconventional		Unconventional		Unconventional	N
Different	Different		Different		Different	
Extraordinary	Extraordinary		Extraordinary		Extraordinary	
Unique	Unique		Unique		Unique	
Renewing	Renewing		Renewing		Renewing	
Hunch	Hunch	N	Hunch	P	Hunch	P
to Think	to Think		to Think		to Think	
Thought	Thought		Thought		Thought	
Mind	Mind		Mind		Mind	
Knowledge	Knowledge		Knowledge		Knowledge	
Resourceful	Resourceful		Resourceful		Resourceful	N
Intelligence	Intelligence		Intelligence		Intelligence	
Imagination	Imagination		Imagination	N	Imagination	P
Inspiration	Inspiration		Inspiration		Inspiration	
Spontaneity	Spontaneity		Spontaneity		Spontaneity	N
Happy	Happy		Happy		Happy	
Feeling	Feeling		Feeling		Feeling	
Passion	Passion		Passion		Passion	
Craftsmanship	Craftsmanship	P	Craftsmanship	P	Craftsmanship	N
(to) knit	(to) knit		(to) knit		(to) knit	
(to) craft	(to) craft		(to) craft		(to) craft	
(to) draw	(to) draw		(to) draw		(to) draw	
(to) play	(to) play		(to) play		(to) play	
(to) design	(to) design		(to) design		(to) design	
Skill	Skill	N	Skill	P	Skill	P
Colours	Colours		Colours		Colours	N
Poem	Poem		Poem	N	Poem	P
Art	Art		Art		Art	
Music	Music		Music		Music	
Artistic	Artistic		Artistic		Artistic	
Painter	Painter		Painter		Painter	
Talent	Talent		Talent		Talent	
(to) make	(to) make		(to) make		(to) make	N

Figure 7: Word clusters based on the first 5 eigenvalues in the PCA of the heat map

The more substantial changes are with the 'novelty' and 'intelligence' groups in figure 4. Five words from the 'intelligence' group in figure 4 are maintained in figure 7 together with *hunch* and *resourceful* from the 'novelty' group in figure 7. Five words from the 'novelty' group in figure 4 are maintained in figure 7 together with *planning*, *process*, and *difficult* from the 'intelligence' group in figure 7.

However, despite these changes there seems to be a substantial overlap in the cluster structure obtained with HCA and PCA. The difference results from the fact that the PCA takes the overall structure of the heat map into account. The clusters as presented in figure 4 and figure 7 could be seen as a semantic map of words related to each other and, as clusters, related to creativity. This map could be used as a basis for the evaluation of creativity.

Semantic map as a basis for evaluation

The literature provides several characteristic of creativity that could be used to evaluate processes or products of computational creativity. As outlined in the introduction these include terms like novel (novelty), original, unique, useful, value, typicality, quality, unexpectedness, surprise, skill, imagination or appreciation.

Many of these are found in the semantic map (figure 4, 7) as well. These include *novel* (*novelty*), *original*, *unique*, *skill*, and *imagination*. Other words are related to words in the semantic map. For example, unexpectedness or surprise are related to *unconventional* and *extraordinary*. The fact that words used in the literature are also found in the semantic map based on empirical investigations underscores their relation with creativity and justifies their use in assessing creativity.

However, some words reported in the literature are notably absent in the semantic map. One of those is the word 'useful'. Although often referred to as a characteristic of creativity (Amabile, 1996; Hennessey and Amabile, 2010; Fink and Benedek, 2014), it is not found in the semantic map. This raises the question of whether humans would qualify useful as related to creativity, and thus as a dimension on which creativity could or should be evaluated.

Because 'useful' did not emerge in our word association studies, we could not investigate its relation with the other terms in the card sorting study. But in a follow up study we will include 'useful' as an item to study its relation to other words related to creativity and to 'creativity' itself in a card sorting study. The outcome will enhance our insight in the way useful and creativity are related as seen by human subjects (instead of by assumption or definition).

One reason of why useful was not included may have resulted from the fact that we asked for terms associated with creativity without any further instruction or direction. It might be that when more specific instructions are given, for example to relate terms to creativity in a particular task or domain, terms like useful might appear.

Hence, another venue of research is to investigate semantic maps related to creativity within specific domains (e.g., music, poetry, architecture), to see if differences between these maps are found. If so, that would argue for more specific forms of evaluation to be used for these domains.

Yet another venue of research is to investigate whether semantic maps (whether or not related to specific domains) also differ between languages. In our association studies (but not the card sorting task itself) we used both Dutch and German native speakers. We could not find significant differences between the two. But this could be related to the similarity between both languages.

The main clusters as presented in figures 4 and 7 could be used to develop rating scales for evaluating the creativity of artificial systems and humans. All of the terms in a cluster could be used as dimensions on which creativity is rated, each one as an example of the main cluster to which it belongs. An analysis of the ratings in terms of the cluster structure could then be related to the clusters found in the semantic map. That is, if the clusters in the semantic map reflect the notions that humans have about creativity, they would also determine the way they evaluate creativity. In that case, evaluations using terms within a cluster would be related to each other and between cluster evaluations would reflect the between cluster structure in the map.

This procedure could also be used for the more domain specific semantic maps, if they are found. In that case, these maps could be used for the evaluation of domain specific forms of creativity and the results of the evaluation could be compared with the structure of the maps.

When more semantic maps are investigated a more complete structure of the semantic relations with creativity will emerge. By comparing this with evaluations of creative processes and products (both computational and human) we will develop a more complete picture of how semantic relations with creativity influence the evaluation of creativity.

The empirically derived semantic maps related to creativity could also be used to develop and evaluate experimental paradigms for investigating the neural basis of creativity. This might begin to unravel the diverse and sometimes apparently conflicting results obtained in the neuroscientific research of creativity (Arden et al., 2010; Dietrich and Kanso, 2010; Sawyer, 2011; Fink and Benedek, 2014).

Effective use of semantic map in evaluation

To use the concepts in the semantic map as tools for evaluation we need to develop and test rating scales based on these concepts. Here, a number of considerations play a role and should be addressed.

The first one is the number of rating scales that can be used effectively. Using all concepts in the map would result in a large set of scales that could be ineffective. We can study this by using the rating scales based on these concepts in pilot evaluations and compare the scales using factor analysis. In this way we can investigate again whether concepts from the same cluster are used in the same way in an evaluation. If so, these rating scales could then be used as alternatives between evaluations. Or they could be used as alternatives within an evaluation (between or within subjects).

The second one concerns the subjects that would perform an evaluation. One option is to use experts in a given domain. Another option is to use the users of a domain in an evaluation. Here, given that the subjects in our studies were students, one can think of creative domains like visual art in gaming (and movies), dance music (and other forms of 'pop' music) and the use of new media like YouTube. Students certainly are involved here as users, and users to a large extent determine success in these domains and thus the way in which these domains develop. It is too simple to argue that only experts determine how forms of creativity develop. Users play a substantial role in that too (as they have also done in the past).

Given a set of rating scales, we can also compare evaluations by experts with that of users. An interesting topic of research here is whether experts in a domain would have a different conceptual structure related to creativity compared to users or whether they would have a similar conceptual structure (as in the semantic map) but would use it differently in an evaluation. This could consist of a different factorization of the rating scales with evaluations performed by experts compared to users.

Conclusions

An empirical basis for the evaluation of creativity is needed because evaluations, as conducted by human raters, are empirical investigations. Hence, the assumptions underlying these investigations, such as the rating scales used, need to be validated. We presented a semantic map of terms related to creativity based on human association and card sorting studies. The semantic map as presented here can be further developed by investigating domain specific aspects of terms related to creativity and the use of other terms often reported as related to creativity in the literature.

To derive the semantic map in the card sorting study, we augmented the words based on our human association studies with words reported in the literature that were based on a statistical analysis. Interestingly, there is an overlap in the set of words formed by the two methods, but there are also some differences. Further investigations could reveal how these methods are related and if they are both needed (as complements) to arrive at more objective procedures for the evaluation of computational (and human) creativity.

Acknowledgements

We thank Saskia Hartmann and Janina Roppelt for their assistance. The work presented here was funded by the project ConCreTe. The project ConCreTe acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET Grant Number 611733.

References

Amabile, T. M. (1996). Creativity and innovation in organizations (Vol. 5). Boston: Harvard Business School.

Arden, R., Chavez, R. S., Grazioplene, R. and Junga, R. E. (2010). Neuroimaging creativity: a psychometric review. Behavioural Brain Research, 214:143156.

Coxon, A. P. M. (1999). Sorting data: Collection and analysis. London, United Kingdom: Sage Publications.

Cardoso A, Veale T, Wiggins GA. (2009). Converging on the divergent: the history (and future) of the international joint workshops in computational creativity. AI Mag, 30(3):15–22.

Colton S. (2008). Creativity versus the perception of creativity in computational systems. In: Proceedings of AAAI symposium on creative systems, p. 14–20.

Dietrich, A. and Kanso, R. (2010). A review of EEG, ERP, and neuroimaging studies of creativity and insight. Psychological Bulletin, 136:822–848.

Fink, A. & Benedek, M. (2014). EEG alpha power and creative ideation. Neuroscience and Biobehavioral Reviews, 44(100): 111–123.

Grace, K. and Maher, M.L. (2014). What to expect when you're expecting: The role of unexpectedness in computationally evaluating creativity, in Proceedings of ICCC2014. http://computationalcreativity.net/iccc2014/proceedings/

Harloff, J., & Coxon, A. P. M. (2006). How to Sort. A Short Guide on Sorting Investigations. http://www.methodofsorting.com/HowToSort1-1_english.pdf (GNU Documentation License).

Hennessey, B.A. & Amabile, T.M. (2010). Creativity. Annual Review of Psychology, 61, 569-598.

Jackson, J. E. (1991). A user's guide to principal components. New York: Wiley.

James, W. (1890). The Principles of Psychology. New York: Henry Holt, Vol. 1, pp. 403–404.

Jolliffe, I. T. (1986). Principal component analysis. New York: Springer.

Jordanous, A. (2012a). Evaluating Computational Creativity: A Standardised Procedure for Evaluating Creative Systems and its Application. Ph.D. Dissertation, University of Sussex, Brighton, UK.

Jordanous, A. (2012b). A standardised procedure for evaluating creative systems: Computational creativity evaluation based on what it is to be creative. Cognitive Computation, 4(3), 246-279.

Jordanous, A. (2014). Stepping Back to Progress Forwards: Setting Standards for Meta-Evaluation of Computational Creativity. In S. Colton, D. Ventura, N. Lavrač and M. Cook (Eds.) Proceedings of the Fifth International Conference on Computational Creativity ICCC-2014, June 10-13, 2014, Ljubljana, Slovenia (pp. 129-136).

Landauer, T. K. and Dumais, S. T. (1997). A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. Psychological Review, 104, 211-240.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. Behavior Research Methods, Instruments, & Computers, 36(3), 402-407.

Ritchie, G. (2007). Some empirical criteria for attributing creativity to a computer program. Minds and Machines, 17(1), 67-99.

Salmoni, A. (2012). Open card sort analysis 101. http://www.uxbooth.com/articles/open-card-sort-analysis-101/

Sawyer, K. (2011). The cognitive neuroscience of creativity: A critical review. Creativity Research Journal, 23(2):137–154.